From:	Benjamin Shorr
To:	Eric Blischke/R10/USEPA/US@EPA; Chip Humphrey/R10/USEPA/US@EPA
Cc:	Robert Gensemer; Carrie Smith; Robert Neely; Ron Gouguet; Jim Koloszar
Subject:	Rd. 2 Data Review- initial spatial analyses
Date:	12/20/2006 11:04 AM

Bob & Eric et al-

I've been going through the contaminant and spatial data and coming up with a methodology/process for querying and summarizing spatially to meet the needs of inputs to HH, ER, BSAF and the other analyses, and mapping/graphing. I've created some base GIS layers that we can uses to summarize/assign location to contaminant data (River Miles & Fate and Transport segments) and have come up with a couple of observations/questions:

1. The reference value table should probably be in the same units as the database (Query Manager) with a clear indication of what guideline or value was chosen based on the priority preference. This will help with identifying the sources in tables/graphs/figures. Additionally, the Chem names should be translated into the Chemcodes in Query Manager- this should help with ensuring consistency between sed/tissue/bioassay & water data & using a look-up table.

2. Statistics: For summarized data- fate and transport segments, River Miles, nearshore receptor habitat etc. I have explored a bit how best to calculate 95% UCL's and perhaps UPLs (using surface sediment as a test case) for 8 metals, Total PAH, PCBs, DDT, Dieldrin. Generally, these contaminants are distributed log-normally (entire site). We should discuss the best and most appropriate way to incorporate/present UCLs/UPLs. Generation of the following statistics for the sub-areas summation is a standard part of the methodology: Min, Max, Count, Mean, SD and Variance.
I've also found that generating a master contaminant data query from Query Manager has some limitations in the GIS because of the -999 entry for non-tested analytes at a station. This just means that folks doing mapping & analysis need to coordinate on what queries should be used for what pieces.

3. Non-detects or below detection limit: It's important to understand how Query Manager queries handle these selections- and how the inputs for the different analyses should be created. We should discuss this.

4. Inclusion/Exclusion of areas like GASCO, T4, McCormick & Baxters: how should data that is in these areas be handled for this data review? Temporally, what data should be used for analysis and presentation for this Rd. 2 Data Review?

I've posted some examples of mapping and Excel graphs (with data tables) for discussion. There may be errors- I've been using these as a test to explore the data and the process.

On our FTP site at:
ftp://ftpuser:ftporr123@ftp.orr.noaa.gov/private/CPRD/PortlandHarbor/Exports/2006_1220_Rd2DataReview/

Maps: 3 series maps showing PAH's summarized by River Miles, AOPC's, and Fate and Transport Segments. 1 map of the River Miles with an explanation of areas.
Excel Table & Graphs: Total PAHs graphed by River Mile and side, Total PAHs in clams (these in particular are not perfect)
Word Doc: Screenshots of selected analytes in surface sediment Histograms & QQPlots
Metadata: Query Manager auto-documentation

I'm sure there is more, but these are initial observations after running through the data a bit, hopefully we can discuss and begin moving forward systematically.

Thanks,

Ben

--
Benjamin Shorr
NOAA National Ocean Service
Assessment and Restoration Division
Physical Scientist, GIS Developer/Analyst
7600 Sand Point Way NE
Seattle, WA 98115

(v) 206.526.4654 (f) 206.526.6865
benjamin.shorr@noaa.gov
http://response.restoration.noaa.gov/orr_about.php